

Approaching nonsmooth nonconvex optimization problems through first order dynamical systems with hidden acceleration and Hessian driven damping terms

Radu Ioan Bot^{*}

Ernö Robert Csetnek[†]

October 5, 2016

Dedicated to the memory of Jon Borwein, who was so inspiring and motivating

Abstract. In this paper we carry out an asymptotic analysis of the proximal-gradient dynamical system

$$\begin{cases} \dot{x}(t) + x(t) = \text{prox}_{\gamma f} [x(t) - \gamma \nabla \Phi(x(t)) - ax(t) - by(t)], \\ \dot{y}(t) + ax(t) + by(t) = 0 \end{cases}$$

where f is a proper, convex and lower semicontinuous function, Φ a possibly nonconvex smooth function and γ, a and b are positive real numbers. We show that the generated trajectories approach the set of critical points of $f + \Phi$, here understood as zeros of its limiting subdifferential, under the premise that a regularization of this sum function satisfies the Kurdyka-Łojasiewicz property. We also establish convergence rates for the trajectories, formulated in terms of the Łojasiewicz exponent of the considered regularization function.

Key Words. dynamical systems, Lyapunov analysis, nonsmooth optimization, limiting subdifferential, Kurdyka-Łojasiewicz property

AMS subject classification. 34G25, 47J25, 47H05, 90C26, 90C30, 65K10

1 Introduction

We begin with a short literature review that serves as motivation for the research conducted in this paper.

The Newton-like dynamical system

$$\ddot{x}(t) + \lambda \dot{x}(t) + \gamma \nabla^2 \Phi(x(t))(\dot{x}(t)) + \nabla \Phi(x(t)) = 0 \quad (1)$$

has been investigated by Alvarez, Attouch, Bolte and Redont in [6] in the context of asymptotically approaching the minimizers of the optimization problem

$$\inf_{x \in \mathbb{R}^n} \Phi(x), \quad (2)$$

for Φ a smooth \mathcal{C}^2 function and λ and γ positive numbers. System (1) is a second order system both in time, due to the presence of the acceleration term $\ddot{x}(t)$, which is associated to inertial effects, and in

^{*}University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, email: radu.bot@univie.ac.at. Research partially supported by FWF (Austrian Science Fund), project I 2419-N32.

[†]University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, email: ernoe.robert.csetnek@univie.ac.at. Research supported by FWF (Austrian Science Fund), Lise Meitner Programme, project M 1682-N25.

space, due to presence of the Hessian $\nabla^2\Phi(x(t))$. The trajectories generated by (1) have been proved to converge to a critical point Φ , when this function is analytic, and to a minimizer of Φ , when it is convex. Dynamical systems of type (1) are of large interest, as they occur in different applications in fields like optimization, mechanics, control theory and PDE theory (see [6, 7, 13, 18–20]).

The authors of [6] have also pointed out the surprising fact that the dynamical system (1) can be viewed as a first order dynamical system with no occurrence of the Hessian. More precisely, it has been shown that (1) is equivalent to

$$\begin{cases} \dot{x}(t) + \gamma \nabla \Phi(x(t)) + ax(t) + by(t) = 0, \\ \dot{y}(t) + ax(t) + by(t) = 0 \end{cases} \quad (3)$$

where $a := \lambda - \frac{1}{\gamma}$ and $b := \frac{1}{\gamma}$. The obvious advantage of (3) comes from the fact that for its asymptotic analysis no second order information on the smooth function Φ is needed. We refer to [6, 19] for applications and other arguments in favor of this reformulation of (1).

On the other hand, in order to asymptotically approach the minimizers of constrained optimization problems of the form

$$\inf_{x \in C} \Phi(x), \quad (4)$$

where $C \subseteq \mathbb{R}^n$ is a nonempty, closed, convex set, the following projection-gradient dynamical system has been considered and investigated by Antipin [8] and Bolte [25]

$$\dot{x}(t) + x(t) = \text{proj}_C (x(t) - \gamma \nabla \Phi(x(t))). \quad (5)$$

Here, $\text{proj}_C : \mathbb{R}^n \rightarrow C$ denotes the *projection operator* onto the set C .

These being given, the following combination of the systems (3) and (5)

$$\begin{cases} \dot{x}(t) + x(t) = \text{proj}_C [x(t) - \gamma \nabla \Phi(x(t)) - ax(t) - by(t)] \\ \dot{y}(t) + ax(t) + by(t) = 0 \end{cases} \quad (6)$$

has been proposed in [6], for a, b and γ positive numbers, in order to asymptotically approach the minimizers of the constrained optimization problem (4) in the hypothesis that the objective function Φ is convex.

Proximal-gradient dynamical systems, which are generalizations of (5), have been recently considered by Abbas and Attouch in [1, Section 5.2] in the full convex setting. Implicit dynamical systems related to both optimization problems and monotone inclusions have been considered in the literature also by Attouch and Svaiter in [21], Attouch, Abbas and Svaiter in [2] and Attouch, Alvarez and Svaiter in [11]. These investigations have been continued and extended in [22, 32, 34–36].

In the last years the interest in approaching the solvability of nonconvex optimization problems from continuous and discrete perspective is continuously increasing (see [12, 14, 15, 29, 31, 37, 39, 40, 43, 47]). Following this tendency, we investigate in this paper the optimization problem

$$\inf_{x \in \mathbb{R}^n} (f(x) + \Phi(x)), \quad (7)$$

where f is a (possibly nonsmooth) proper, convex and lower semicontinuous function and Φ a (possibly nonconvex) smooth function. More precisely, in this paper we investigate the convergence of the trajectories generated by the proximal-gradient dynamical system

$$\begin{cases} \dot{x}(t) + x(t) = \text{prox}_{\gamma f} [x(t) - \gamma \nabla \Phi(x(t)) - ax(t) - by(t)], \\ \dot{y}(t) + ax(t) + by(t) = 0 \end{cases} \quad (8)$$

where a, b and γ are positive real numbers and

$$\text{prox}_{\gamma f} : \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad \text{prox}_{\gamma f}(y) = \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ f(u) + \frac{1}{2\gamma} \|u - y\|^2 \right\},$$

denotes the *proximal point operator* of γf , to a critical point of $f + \Phi$, here understood as a zero of its limiting subdifferential. To this end we assume that a regularization of the objective function satisfies the *Kurdyka-Łojasiewicz property*; in other words, it is a *KL function*. The convergence analysis relies on methods and concepts of real algebraic geometry introduced by Łojasiewicz [45] and Kurdyka [44] and later developed in the nonsmooth setting by Attouch, Bolte and Svaiter [15] and Bolte, Sabach and Teboulle [29].

In the convergence analysis we use three main ingredients: (1) we prove a Lyapunov-type property, expressed as a sufficient decrease of a regularization of the objective function along the trajectories, (2) we show the existence of a subgradient lower bound for the trajectories and, finally, (3) we derive convergence by making use of the Kurdyka-Łojasiewicz property of the objective function (for a similar approach in the continuous case see [6] and in the discrete setting see [15, 29]). Furthermore, we obtain convergence rates for the trajectories expressed in terms of the Łojasiewicz exponent of the regularized objective function.

2 Preliminaries

We recall some notions and results which are needed throughout the paper. We consider on \mathbb{R}^n the Euclidean scalar product and the corresponding norm denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively.

The *domain* of the function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is defined by $\text{dom } f = \{x \in \mathbb{R}^n : f(x) < +\infty\}$. We say that f is *proper*, if $\text{dom } f \neq \emptyset$. For the following generalized subdifferential notions and their basic properties we refer to [30, 46, 48]. Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. The *Fréchet (viscosity) subdifferential* of f at $x \in \text{dom } f$ is the set

$$\hat{\partial}f(x) = \left\{ v \in \mathbb{R}^n : \liminf_{y \rightarrow x} \frac{f(y) - f(x) - \langle v, y - x \rangle}{\|y - x\|} \geq 0 \right\}.$$

For $x \notin \text{dom } f$, one sets $\hat{\partial}f(x) := \emptyset$. The *limiting (Mordukhovich) subdifferential* is defined at $x \in \text{dom } f$ by

$$\partial f(x) = \{v \in \mathbb{R}^n : \exists x_k \rightarrow x, f(x_k) \rightarrow f(x) \text{ and } \exists v_k \in \hat{\partial}f(x_k), v_k \rightarrow v \text{ as } k \rightarrow +\infty\},$$

while for $x \notin \text{dom } f$, one takes $\partial f(x) := \emptyset$. Therefore $\hat{\partial}f(x) \subseteq \partial f(x)$ for each $x \in \mathbb{R}^n$.

When f is convex, these subdifferential notions coincide with the *convex subdifferential*, thus $\hat{\partial}f(x) = \partial f(x) = \{v \in \mathbb{R}^n : f(y) \geq f(x) + \langle v, y - x \rangle \ \forall y \in \mathbb{R}^n\}$ for all $x \in \mathbb{R}^n$.

The following *closedness criterion* of the graph of the limiting subdifferential will be used in the convergence analysis: if $(x_k)_{k \in \mathbb{N}}$ and $(v_k)_{k \in \mathbb{N}}$ are sequences in \mathbb{R}^n such that $v_k \in \partial f(x_k)$ for all $k \in \mathbb{N}$, $(x_k, v_k) \rightarrow (x, v)$ and $f(x_k) \rightarrow f(x)$ as $k \rightarrow +\infty$, then $v \in \partial f(x)$.

The Fermat rule reads in this nonsmooth setting as follows: if $x \in \mathbb{R}^n$ is a local minimizer of f , then $0 \in \partial f(x)$. We denote by

$$\text{crit}(f) = \{x \in \mathbb{R}^n : 0 \in \partial f(x)\}$$

the set of *(limiting)-critical points* of f .

When f is continuously differentiable around $x \in \mathbb{R}^n$ we have $\partial f(x) = \{\nabla f(x)\}$. We will also make use of the following subdifferential sum rule: if $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is proper and lower semicontinuous and $h : \mathbb{R}^n \rightarrow \mathbb{R}$ is a continuously differentiable function, then $\partial(f + h)(x) = \partial f(x) + \nabla h(x)$ for all $x \in \mathbb{R}^n$.

A crucial role in the asymptotic analysis of the dynamical system (8) is played by the class of functions satisfying the *Kurdyka-Łojasiewicz property*. For $\eta \in (0, +\infty]$, we denote by Θ_η the class of concave and continuous functions $\varphi : [0, \eta] \rightarrow [0, +\infty)$ such that $\varphi(0) = 0$, φ is continuously differentiable on $(0, \eta)$, continuous at 0 and $\varphi'(s) > 0$ for all $s \in (0, \eta)$. In the following definition (see [14, 29]) we use also the *distance function* to a set, defined for $A \subseteq \mathbb{R}^n$ as $\text{dist}(x, A) = \inf_{y \in A} \|x - y\|$ for all $x \in \mathbb{R}^n$.

Definition 1 (*Kurdyka-Łojasiewicz property*) Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper and lower semi-continuous function. We say that f satisfies the *Kurdyka-Łojasiewicz (KL) property* at $\bar{x} \in \text{dom } \partial f = \{x \in \mathbb{R}^n : \partial f(x) \neq \emptyset\}$, if there exist $\eta \in (0, +\infty]$, a neighborhood U of \bar{x} and a function $\varphi \in \Theta_\eta$ such that for all x in the intersection

$$U \cap \{x \in \mathbb{R}^n : f(\bar{x}) < f(x) < f(\bar{x}) + \eta\}$$

the following inequality holds

$$\varphi'(f(x) - f(\bar{x})) \text{dist}(0, \partial f(x)) \geq 1.$$

If f satisfies the KL property at each point in $\text{dom } \partial f$, then f is called *KL function*.

The origins of this notion go back to the pioneering work of Łojasiewicz [45], where it is proved that for a real-analytic function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and a critical point $\bar{x} \in \mathbb{R}^n$ (that is $\nabla f(\bar{x}) = 0$), there exists $\theta \in [1/2, 1)$ such that the function $|f - f(\bar{x})|^\theta \|\nabla f\|^{-1}$ is bounded around \bar{x} . This corresponds to the situation when $\varphi(s) = Cs^{1-\theta}$, where $C > 0$. The result of Łojasiewicz allows the interpretation of the KL property as a re-parametrization of the function values in order to avoid flatness around the critical points. Kurdyka [44] extended this property to differentiable functions definable in o-minimal structures. Further extensions to the nonsmooth setting can be found in [14, 26–28].

One of the remarkable properties of the KL functions is their ubiquity in applications (see [29]). To the class of KL functions belong semi-algebraic, real sub-analytic, semiconvex, uniformly convex and convex functions satisfying a growth condition. We refer the reader to [12, 14, 15, 26–29] and the references therein for more on KL functions and illustrating examples.

In the analysis below the following uniform KL property given in [29, Lemma 6] will be used.

Lemma 1 Let $\Omega \subseteq \mathbb{R}^n$ be a compact set and let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. Assume that f is constant on Ω and that it satisfies the KL property at each point of Ω . Then there exist $\varepsilon, \eta > 0$ and $\varphi \in \Theta_\eta$ such that for all $\bar{x} \in \Omega$ and all x in the intersection

$$\{x \in \mathbb{R}^n : \text{dist}(x, \Omega) < \varepsilon\} \cap \{x \in \mathbb{R}^n : f(\bar{x}) < f(x) < f(\bar{x}) + \eta\} \quad (9)$$

the inequality

$$\varphi'(f(x) - f(\bar{x})) \text{dist}(0, \partial f(x)) \geq 1. \quad (10)$$

holds.

In the following we recall the notion of locally absolutely continuous function and state two of its basic properties.

Definition 2 (see, for instance, [2, 21]) A function $x : [0, +\infty) \rightarrow \mathbb{R}^n$ is said to be locally absolutely continuous, if it is absolutely continuous on every interval $[0, T]$, where $T > 0$.

Remark 2 (a) An absolutely continuous function is differentiable almost everywhere, its derivative coincides with its distributional derivative almost everywhere and one can recover the function from its derivative $\dot{x} = y$ by integration.

(b) If $x : [0, T] \rightarrow \mathbb{R}^n$ is absolutely continuous for $T > 0$ and $B : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is L -Lipschitz continuous for $L \geq 0$, then the function $z = B \circ x$ is absolutely continuous, too. Moreover, z is differentiable almost everywhere on $[0, T]$ and the inequality $\|\dot{z}(t)\| \leq L\|\dot{x}(t)\|$ holds for almost every $t \in [0, T]$.

The following two results, which can be interpreted as continuous versions of the quasi-Fejér monotonicity for sequences, will play an important role in the asymptotic analysis of the trajectories of the dynamical system investigated in this paper. For their proofs we refer the reader to [2, Lemma 5.1] and [2, Lemma 5.2], respectively.

Lemma 3 Suppose that $F : [0, +\infty) \rightarrow \mathbb{R}$ is locally absolutely continuous and bounded from below and that there exists $G \in L^1([0, +\infty))$ such that for almost every $t \in [0, +\infty)$

$$\frac{d}{dt}F(t) \leq G(t).$$

Then there exists $\lim_{t \rightarrow \infty} F(t) \in \mathbb{R}$.

Lemma 4 If $1 \leq p < \infty$, $1 \leq r \leq \infty$, $F : [0, +\infty) \rightarrow [0, +\infty)$ is locally absolutely continuous, $F \in L^p([0, +\infty))$, $G : [0, +\infty) \rightarrow \mathbb{R}$, $G \in L^r([0, +\infty))$ and for almost every $t \in [0, +\infty)$

$$\frac{d}{dt}F(t) \leq G(t),$$

then $\lim_{t \rightarrow +\infty} F(t) = 0$.

Further we recall a differentiability result that involves the composition of convex functions with absolutely continuous trajectories, which is due to Brézis ([38, Lemme 3.3, p. 73]; see also [16, Lemma 3.2]).

Lemma 5 Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper, convex and lower semicontinuous function. Let $x \in L^2([0, T], \mathbb{R}^n)$ be absolutely continuous such that $\dot{x} \in L^2([0, T], \mathbb{R}^n)$ and $x(t) \in \text{dom } f$ for almost every $t \in [0, T]$. Assume that there exists $\xi \in L^2([0, T], \mathbb{R}^n)$ such that $\xi(t) \in \partial f(x(t))$ for almost every $t \in [0, T]$. Then the function $t \mapsto f(x(t))$ is absolutely continuous and for almost every t such that $x(t) \in \text{dom } \partial f$ we have

$$\frac{d}{dt}f(x(t)) = \langle \dot{x}(t), h \rangle \quad \forall h \in \partial f(x(t)).$$

We close this section with the following characterization of the proximal point operator of a proper, convex and lower semicontinuous function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$: for every $\gamma > 0$ it holds (see for example [23])

$$p = \text{prox}_{\gamma f}(x) \text{ if and only if } x \in p + \gamma \partial f(p), \quad (11)$$

where ∂f denotes the convex subdifferential of f .

3 Asymptotic analysis

The dynamical system we investigate in this paper reads

$$\begin{cases} \dot{x}(t) + x(t) = \text{prox}_{\gamma f} [x(t) - \gamma \nabla \Phi(x(t)) - ax(t) - by(t)], \\ \dot{y}(t) + ax(t) + by(t) = 0 \\ x(0) = x_0, y(0) = y_0, \end{cases} \quad (12)$$

where $x_0, y_0 \in \mathbb{R}^n$ and a, b and γ are positive real numbers. We assume that $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is proper, convex and lower semicontinuous, while $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$ is a Fréchet differentiable with L -Lipschitz continuous gradient, for $L > 0$, that is $\|\nabla \Phi(x) - \nabla \Phi(y)\| \leq L\|x - y\|$ for all $x, y \in \mathbb{R}^n$.

The existence and uniqueness of the trajectories generated by (12) can be proved by using the estimates from the proof of Lemma 6 below and by following a classical argument, as in [6, Theorem 7.1].

For the asymptotic analysis, we impose on the parameters involved the following condition:

$$\begin{cases} 2\gamma L(|1 - a| + \gamma L) + |1 - a| + \gamma L + b\gamma L < 1 \\ ab + \frac{a}{2} + \frac{1}{2}a|1 - a| + \frac{1}{2}\gamma aL + \frac{1}{2}\gamma abL < b \end{cases} \quad (13)$$

and notice that the first inequality is fulfilled for an arbitrary $b > 0$, if $a \in (0, 2)$ and $\gamma > 0$ are chosen small enough, while the second one holds for $a > 0$ small enough.

3.1 Convergence of the trajectories

We begin with the proof of a decrease property for a regularization of the objective function along the trajectories.

Lemma 6 *Suppose that $f + \Phi$ is bounded from below and the parameters a, b, γ and L satisfy (13). For $x_0, y_0 \in \mathbb{R}^n$, let $(x, y) \in C^1([0, +\infty), \mathbb{R}^n) \times C^2([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (12). Then the following statements are true:*

(a) $\frac{d}{dt} \left[(f + \Phi)(\dot{x}(t) + x(t)) + \frac{1}{2\gamma} \|\dot{x}(t)\|^2 + \frac{1}{2\gamma a} \|ax(t) + by(t)\|^2 \right] \leq -M_1 \|\dot{x}(t)\|^2 - M_2 \|\dot{y}(t)\|^2$ for almost every $t \geq 0$, where

$$M_1 := \frac{1}{2\gamma} - L(|1 - a| + \gamma L) - \frac{1}{2\gamma}|1 - a| - \frac{1}{2}L - \frac{1}{2}bL > 0$$

and

$$M_2 := \frac{b}{\gamma a} - \frac{b}{\gamma} - \frac{1}{2\gamma} - \frac{1}{2\gamma}|1 - a| - \frac{1}{2}L - \frac{1}{2}bL > 0;$$

(b) $\dot{x}, \dot{y}, ax + by \in L^2([0, +\infty); \mathbb{R}^n)$ and $\lim_{t \rightarrow +\infty} \dot{x}(t) = \lim_{t \rightarrow +\infty} \dot{y}(t) = \lim_{t \rightarrow +\infty} (ax(t) + by(t)) = 0$;

(c) $\exists \lim_{t \rightarrow +\infty} (f + \Phi)(\dot{x}(t) + x(t)) \in \mathbb{R}$.

Proof. Define $z : [0, +\infty) \rightarrow \mathbb{R}^n$ by

$$z(t) = \text{prox}_{\gamma f} [x(t) - \gamma \nabla \Phi(x(t)) - ax(t) - by(t)]. \quad (14)$$

Since $\text{prox}_{\gamma f}$ is nonexpansive (that is 1-Lipschitz continuous), in view of Remark 2(b), z is locally absolutely continuous. From the Lipschitz continuity of $\nabla \Phi$ we obtain

$$\|z(t) - z(s)\| \leq (|1 - a| + \gamma L)\|x(t) - x(s)\| + b\|y(t) - y(s)\| \quad \forall t, s \geq 0,$$

hence, for almost every $t \geq 0$,

$$\|\dot{z}(t)\| \leq (|1 - a| + \gamma L)\|\dot{x}(t)\| + b\|\dot{y}(t)\|. \quad (15)$$

Since

$$\dot{x}(t) + x(t) = z(t) \quad \forall t \geq 0, \quad (16)$$

it follows that \dot{x} is locally absolutely continuous, hence \ddot{x} exists almost everywhere on $[0, +\infty)$ and for almost every $t \geq 0$ it holds

$$\|\ddot{x}(t)\| \leq (1 + |1 - a| + \gamma L)\|\dot{x}(t)\| + b\|\dot{y}(t)\|. \quad (17)$$

We fix an arbitrary $T > 0$. From the characterization (11) of the proximal point operator we have

$$-\frac{1}{\gamma}\dot{x}(t) - \frac{a}{\gamma}x(t) - \frac{b}{\gamma}y(t) - \nabla \Phi(x(t)) \in \partial f(\dot{x}(t) + x(t)) \quad \forall t \in [0, +\infty). \quad (18)$$

Due to the continuity properties of the trajectories and their derivatives on $[0, T]$, (17) and the Lipschitz continuity of $\nabla \Phi$, we have

$$x, \dot{x}, \dot{y}, \ddot{x}, \nabla \Phi(x) \in L^2([0, T]; \mathbb{R}^n).$$

Applying Lemma 5 we obtain that the function $t \mapsto f(\dot{x}(t) + x(t))$ is absolutely continuous and

$$\frac{d}{dt} f(\dot{x}(t) + x(t)) = \left\langle -\frac{1}{\gamma}\dot{x}(t) - \frac{a}{\gamma}x(t) - \frac{b}{\gamma}y(t) - \nabla \Phi(x(t)), \ddot{x}(t) + \dot{x}(t) \right\rangle$$

for almost every $t \in [0, T]$. Moreover, it holds

$$\frac{d}{dt}\Phi(\dot{x}(t) + x(t)) = \langle \nabla\Phi(\dot{x}(t) + x(t)), \ddot{x}(t) + \dot{x}(t) \rangle$$

for almost every $t \in [0, T]$. Summing up the last two equalities and by taking into account (12), we obtain

$$\begin{aligned} \frac{d}{dt}(f + \Phi)(\dot{x}(t) + x(t)) &= -\frac{1}{2\gamma}\frac{d}{dt}(\|\dot{x}(t)\|^2) - \frac{1}{\gamma}\|\dot{x}(t)\|^2 - \frac{1}{\gamma}\langle ax(t) + by(t), \ddot{x}(t) + \dot{x}(t) \rangle \\ &\quad + \langle \nabla\Phi(\dot{x}(t) + x(t)) - \nabla\Phi(x(t)), \ddot{x}(t) + \dot{x}(t) \rangle \\ &= -\frac{1}{2\gamma}\frac{d}{dt}(\|\dot{x}(t)\|^2) - \frac{1}{\gamma}\|\dot{x}(t)\|^2 + \frac{1}{\gamma}\langle \dot{y}(t), \ddot{x}(t) + \dot{x}(t) \rangle \\ &\quad + \langle \nabla\Phi(\dot{x}(t) + x(t)) - \nabla\Phi(x(t)), \ddot{x}(t) + \dot{x}(t) \rangle \end{aligned} \quad (19)$$

for almost every $t \in [0, T]$. Further, due to (12) we have

$$\begin{aligned} \frac{d}{dt}\left(\frac{1}{2}\|ax(t) + by(t)\|^2\right) &= \langle ax(t) + by(t), a\dot{x}(t) + b\dot{y}(t) \rangle \\ &= -a\langle \dot{x}(t), \dot{y}(t) \rangle - b\|\dot{y}(t)\|^2. \end{aligned}$$

Substituting the term $\langle \dot{x}(t), \dot{y}(t) \rangle$ from the last relation into (19) we get

$$\begin{aligned} \frac{d}{dt}(f + \Phi)(\dot{x}(t) + x(t)) &= -\frac{1}{2\gamma}\frac{d}{dt}(\|\dot{x}(t)\|^2) - \frac{1}{\gamma}\|\dot{x}(t)\|^2 \\ &\quad - \frac{1}{\gamma a}\frac{d}{dt}\left(\frac{1}{2}\|ax(t) + by(t)\|^2\right) - \frac{b}{\gamma a}\|\dot{y}(t)\|^2 \\ &\quad + \frac{1}{\gamma}\langle \dot{y}(t), \ddot{x}(t) \rangle + \langle \nabla\Phi(\dot{x}(t) + x(t)) - \nabla\Phi(x(t)), \ddot{x}(t) + \dot{x}(t) \rangle \\ &\leq -\frac{1}{2\gamma}\frac{d}{dt}(\|\dot{x}(t)\|^2) - \frac{1}{\gamma}\|\dot{x}(t)\|^2 - \frac{b}{\gamma a}\|\dot{y}(t)\|^2 \\ &\quad - \frac{1}{\gamma a}\frac{d}{dt}\left(\frac{1}{2}\|ax(t) + by(t)\|^2\right) + \frac{1}{\gamma}(1 + |1 - a| + \gamma L)\|\dot{x}(t)\| \cdot \|\dot{y}(t)\| \\ &\quad + \frac{b}{\gamma}\|\dot{y}(t)\|^2 + L\|\dot{x}(t)\| \cdot \|\ddot{x}(t) + \dot{x}(t)\| \end{aligned}$$

for almost every $t \in [0, T]$. Noticing that

$$\|\ddot{x}(t) + \dot{x}(t)\| = \|\dot{z}(t)\|$$

and by taking into account (15), we derive

$$\begin{aligned} \frac{d}{dt}(f + \Phi)(\dot{x}(t) + x(t)) &\leq -\frac{1}{2\gamma}\frac{d}{dt}(\|\dot{x}(t)\|^2) - \frac{1}{\gamma a}\frac{d}{dt}\left(\frac{1}{2}\|ax(t) + by(t)\|^2\right) \\ &\quad - \left(\frac{1}{\gamma} - L(|1 - a| + \gamma L)\right)\|\dot{x}(t)\|^2 - \left(\frac{b}{\gamma a} - \frac{b}{\gamma}\right)\|\dot{y}(t)\|^2 \\ &\quad + \frac{1}{\gamma}(1 + |1 - a| + \gamma L + \gamma bL)\|\dot{x}(t)\| \cdot \|\dot{y}(t)\| \end{aligned}$$

for almost every $t \in [0, T]$. Finally, by using the inequality $\|\dot{x}(t)\| \cdot \|\dot{y}(t)\| \leq \frac{1}{2}\|\dot{x}(t)\|^2 + \frac{1}{2}\|\dot{y}(t)\|^2$ and by taking into account the definitions of M_1 and M_2 , we conclude that (a) holds.

(b) By integration we get

$$\begin{aligned} & (f + \Phi)(\dot{x}(T) + x(T)) + \frac{1}{2\gamma}\|\dot{x}(T)\|^2 + \frac{1}{2\gamma a}\|ax(T) + by(T)\|^2 + M_1 \int_0^T \|\dot{x}(t)\|^2 dt \\ & + M_2 \int_0^T \|\dot{y}(t)\|^2 dt \leq (f + \Phi)(\dot{x}(0) + x_0) + \frac{1}{2\gamma}\|\dot{x}(0)\|^2 + \frac{1}{2\gamma a}\|ax_0 + by_0\|^2. \end{aligned} \quad (20)$$

Since $f + \Phi$ is bounded from below and by taking into account that $T > 0$ has been arbitrarily chosen, we obtain

$$\dot{x}, \dot{y} \in L^2([0, +\infty); \mathbb{R}^n). \quad (21)$$

Due to (17), this further implies

$$\ddot{x} \in L^2([0, +\infty); \mathbb{R}^n). \quad (22)$$

Furthermore, for almost every $t \in [0, +\infty)$ we have

$$\frac{d}{dt}(\|\dot{x}(t)\|^2) = 2\langle \dot{x}(t), \ddot{x}(t) \rangle \leq \|\dot{x}(t)\|^2 + \|\ddot{x}(t)\|^2.$$

By applying Lemma 4, it follows that $\lim_{t \rightarrow +\infty} \dot{x}(t) = 0$. Moreover, from (12) we get that \ddot{y} exists and $\ddot{y} \in L^2([0, +\infty); \mathbb{R}^n)$ due to (21). The same arguments are used in order to conclude $\lim_{t \rightarrow +\infty} \dot{y}(t) = 0$.

(c) From (a) we get

$$\frac{d}{dt} \left[(f + \Phi)(\dot{x}(t) + x(t)) + \frac{1}{2\gamma}\|\dot{x}(t)\|^2 + \frac{1}{2\gamma a}\|ax(t) + by(t)\|^2 \right] \leq 0$$

for almost every $t \geq 0$. From Lemma 3 it follows that

$$\lim_{t \rightarrow +\infty} \left[(f + \Phi)(\dot{x}(t) + x(t)) + \frac{1}{2\gamma}\|\dot{x}(t)\|^2 + \frac{1}{2\gamma a}\|ax(t) + by(t)\|^2 \right]$$

exists and it is a real number, hence from

$$\lim_{t \rightarrow +\infty} \dot{x}(t) = \lim_{t \rightarrow +\infty} \dot{y}(t) = \lim_{t \rightarrow +\infty} (-ax(t) - by(t)) = 0$$

the conclusion follows. ■

We define the *limit set* of x as

$$\omega(x) = \{\bar{x} \in \mathbb{R}^n : \exists t_k \rightarrow +\infty \text{ such that } x(t_k) \rightarrow \bar{x} \text{ as } k \rightarrow +\infty\}.$$

Lemma 7 *Suppose that $f + \Phi$ is bounded from below and the parameters a, b, γ and L satisfy (13). For $x_0, y_0 \in \mathbb{R}^n$, let $(x, y) \in C^1([0, +\infty), \mathbb{R}^n) \times C^2([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (12). Then*

$$\omega(x) \subseteq \text{crit}(f + \Phi).$$

Proof. Let $\bar{x} \in \omega(x)$ and $t_k \rightarrow +\infty$ be such that $x(t_k) \rightarrow \bar{x}$ as $k \rightarrow +\infty$. From (18) we have

$$\begin{aligned} & -\frac{1}{\gamma}\dot{x}(t_k) - \frac{a}{\gamma}x(t_k) - \frac{b}{\gamma}y(t_k) - \nabla\Phi(x(t_k)) + \nabla\Phi(\dot{x}(t_k) + x(t_k)) \\ & \in \partial f(\dot{x}(t_k) + x(t_k)) + \nabla\Phi(\dot{x}(t_k) + x(t_k)) = \partial(f + \Phi)(\dot{x}(t_k) + x(t_k)) \quad \forall k \in \mathbb{N}. \end{aligned} \quad (23)$$

Lemma 6(b), (12) and the Lipschitz continuity of $\nabla\Phi$ ensure that

$$-\frac{1}{\gamma}\dot{x}(t_k) - \frac{a}{\gamma}x(t_k) - \frac{b}{\gamma}y(t_k) - \nabla\Phi(x(t_k)) + \nabla\Phi(\dot{x}(t_k) + x(t_k)) \rightarrow 0 \text{ as } k \rightarrow +\infty \quad (24)$$

and

$$\dot{x}(t_k) + x(t_k) \rightarrow \bar{x} \text{ as } k \rightarrow +\infty. \quad (25)$$

We claim that

$$\lim_{k \rightarrow +\infty} (f + \Phi)(\dot{x}(t_k) + x(t_k)) = (f + \Phi)(\bar{x}). \quad (26)$$

Indeed, from (25) and the lower semicontinuity of f we get

$$\liminf_{k \rightarrow +\infty} f(\dot{x}(t_k) + x(t_k)) \geq f(\bar{x}). \quad (27)$$

Further, since

$$\begin{aligned} \dot{x}(t_k) + x(t_k) &= \operatorname{argmin}_{u \in \mathbb{R}^n} \left[f(u) + \frac{1}{2\gamma} \|u - (x(t_k) - \gamma \nabla \Phi(x(t_k)) - ax(t_k) - by(t_k))\|^2 \right] \\ &= \operatorname{argmin}_{u \in \mathbb{R}^n} \left[f(u) + \frac{1}{2\gamma} \|u - (x(t_k) - ax(t_k) - by(t_k))\|^2 + \langle u - (x(t_k) - ax(t_k) - by(t_k)), \nabla \Phi(x(t_k)) \rangle \right], \end{aligned}$$

we have the inequality

$$\begin{aligned} &f(\dot{x}(t_k) + x(t_k)) + \frac{1}{2\gamma} \|\dot{x}(t_k) - ax(t_k) - by(t_k)\|^2 + \langle \dot{x}(t_k) - ax(t_k) - by(t_k), \nabla \Phi(x(t_k)) \rangle \\ &\leq f(\bar{x}) + \frac{1}{2\gamma} \|\bar{x} - (x(t_k) - ax(t_k) - by(t_k))\|^2 + \langle \bar{x} - (x(t_k) - ax(t_k) - by(t_k)), \nabla \Phi(x(t_k)) \rangle \quad \forall k \in \mathbb{N}. \end{aligned}$$

Taking in the above inequality the limit as $k \rightarrow +\infty$, we derive by using again Lemma 6(b) that

$$\limsup_{k \rightarrow +\infty} f(\dot{x}(t_k) + x(t_k)) \leq f(\bar{x}),$$

which combined with (27) implies

$$\lim_{k \rightarrow +\infty} f(\dot{x}(t_k) + x(t_k)) = f(\bar{x}).$$

By using (25) and the continuity of Φ we conclude that (26) is true.

Altogether, from (23), (24), (25), (26) and the closedness criteria of the limiting subdifferential we obtain $0 \in \partial(f + \Phi)(\bar{x})$ and the proof is complete. \blacksquare

Lemma 8 *Suppose that $f + \Phi$ is bounded from below and the parameters a, b, γ and L satisfy (13). For $x_0, y_0 \in \mathbb{R}^n$, let $(x, y) \in C^1([0, +\infty), \mathbb{R}^n) \times C^2([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (12). Consider the function*

$$H : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}, \quad H(u, v, w) = (f + \Phi)(u) + \frac{1}{2\gamma} \|u - v\|^2 + \frac{1}{2\gamma a} \|av + bw\|^2.$$

Then the following statements are true:

(H₁) for almost every $t \in [0, +\infty)$ it holds

$$\frac{d}{dt} H(\dot{x}(t) + x(t), x(t), y(t)) \leq -M_1 \|\dot{x}(t)\|^2 - M_2 \|\dot{y}(t)\|^2 \leq 0$$

and

$$\exists \lim_{t \rightarrow +\infty} H(\dot{x}(t) + x(t), x(t), y(t)) \in \mathbb{R};$$

(H₂) when $\zeta : [0, +\infty) \rightarrow \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$ is defined by

$$\zeta(t) := \left(-\nabla\Phi(x(t)) + \nabla\Phi(\dot{x}(t) + x(t)) + \frac{1}{\gamma}\dot{y}(t), -\frac{1}{\gamma}\dot{x}(t) - \frac{1}{\gamma}\dot{y}(t), -\frac{b}{\gamma a}\dot{y}(t) \right),$$

then for every $t \in [0, +\infty)$ it holds

$$\zeta(t) \in \partial H(\dot{x}(t) + x(t), x(t), y(t))$$

and

$$\|\zeta(t)\| \leq \left(\frac{2}{\gamma} + \frac{b}{\gamma a} \right) \|\dot{y}(t)\| + \left(L + \frac{1}{\gamma} \right) \|\dot{x}(t)\|;$$

(H₃) for $\bar{x} \in \omega(x)$ and $t_k \rightarrow +\infty$ such that $x(t_k) \rightarrow \bar{x}$ as $k \rightarrow +\infty$, it holds

$$H(\dot{x}(t_k) + x(t_k), x(t_k), y(t_k)) \rightarrow (f + \Phi)(\bar{x}) = H\left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}\right) \text{ as } k \rightarrow +\infty.$$

Proof. (H₁) follows from Lemma 6. The first statement in (H₂) is a consequence of (18), the equation $\dot{y}(t) + ax(t) + by(t) = 0$ and the fact that

$$\partial H(u, v, w) = \left(\partial(f + \Phi)(u) + \frac{1}{\gamma}(u - v) \right) \times \left\{ \frac{1}{\gamma}(v - u) + \frac{1}{\gamma}(av + bw) \right\} \times \left\{ \frac{b}{\gamma a}(av + bw) \right\} \quad (28)$$

for all $(u, v, w) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$. The second statement in (H₂) is a consequence of the Lipschitz continuity of $\nabla\Phi$. Finally, (H₃) has been shown as intermediate step in the proof of Lemma 7. ■

Lemma 9 Suppose that $f + \Phi$ is bounded from below and the parameters a, b, γ and L satisfy (13). For $x_0, y_0 \in \mathbb{R}^n$, let $(x, y) \in C^1([0, +\infty), \mathbb{R}^n) \times C^2([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (12). Consider the function

$$H : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}, H(u, v, w) = (f + \Phi)(u) + \frac{1}{2\gamma}\|u - v\|^2 + \frac{1}{2\gamma a}\|av + bw\|^2.$$

Suppose that x is bounded. Then the following statements are true:

- (a) $\omega(\dot{x} + x, x, y) \subseteq \text{crit}(H) = \{(u, u, -\frac{a}{b}u) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n : u \in \text{crit}(f + \Phi)\};$
- (b) $\lim_{t \rightarrow +\infty} \text{dist}\left((\dot{x}(t) + x(t), x(t), y(t)), \omega(\dot{x} + x, x, y)\right) = 0;$
- (c) $\omega(\dot{x} + x, x, y)$ is nonempty, compact and connected;
- (d) H is finite and constant on $\omega(\dot{x} + x, x, y)$.

Proof. (a), (b) and (d) are direct consequences of Lemma 6, Lemma 7 and Lemma 8.

Finally, (c) is a classical result from [41]. We also refer the reader to the proof of Theorem 4.1 in [6], where it is shown that the properties of $\omega(x)$ of being nonempty, compact and connected are generic for bounded trajectories fulfilling $\lim_{t \rightarrow +\infty} \dot{x}(t) = 0$. ■

Remark 10 Suppose that a, b, γ and $L > 0$ fulfill the inequality (13) and $f + \Phi$ is coercive, in other words,

$$\lim_{\|u\| \rightarrow +\infty} (f + \Phi)(u) = +\infty.$$

For $x_0, y_0 \in \mathbb{R}^n$, let $(x, y) \in C^1([0, +\infty), \mathbb{R}^n) \times C^2([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (12). Then $f + \Phi$ is bounded from below and x is bounded.

Indeed, since $f + \Phi$ is a proper, lower semicontinuous and coercive function, it follows that $\inf_{u \in \mathbb{R}^n} [f(u) + \Phi(u)]$ is finite and the infimum is attained. Hence $f + \Phi$ is bounded from below. On the other hand, from (20) it follows

$$(f + \Phi)(\dot{x}(T) + x(T)) \leq (f + \Phi)(\dot{x}(0) + x_0) + \frac{1}{2\gamma} \|\dot{x}(0)\|^2 + \frac{1}{2\gamma a} \|ax_0 + by_0\|^2 \quad \forall T \geq 0.$$

Since $f + \Phi$ is coercive, the lower level sets of $f + \Phi$ are bounded, hence the above inequality yields that $\dot{x} + x$ is bounded, which combined with $\lim_{t \rightarrow +\infty} \dot{x}(t) = 0$ delivers the boundedness of x . Notice that in this case y is bounded, too, due to Lemma 6(b) and the equation $\dot{y}(t) + ax(t) + by(t) = 0$.

Now we are in the position to present the first main result of the paper, which concerns the convergence of the trajectories generated by (12).

Theorem 11 *Suppose that $f + \Phi$ is bounded from below and the parameters a, b, γ and L satisfy (13). For $x_0, y_0 \in \mathbb{R}^n$, let $(x, y) \in C^1([0, +\infty), \mathbb{R}^n) \times C^2([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (12). Consider the function*

$$H : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}, \quad H(u, v, w) = (f + \Phi)(u) + \frac{1}{2\gamma} \|u - v\|^2 + \frac{1}{2\gamma a} \|av + bw\|^2.$$

Suppose that x is bounded. Then the following statements are true:

- (a) $\dot{x}, \dot{y}, ax + by \in L^1([0, +\infty); \mathbb{R}^n)$ and $\lim_{t \rightarrow +\infty} \dot{x}(t) = \lim_{t \rightarrow +\infty} \dot{y}(t) = \lim_{t \rightarrow +\infty} (ax(t) + by(t)) = 0$;
- (b) *there exists $\bar{x} \in \text{crit}(f + \Phi)$ such that $\lim_{t \rightarrow +\infty} x(t) = \bar{x}$ and $\lim_{t \rightarrow +\infty} y(t) = -\frac{a}{b}\bar{x}$.*

Proof. According to Lemma 9, we can choose an element $\bar{x} \in \text{crit}(f + \Phi)$ such that $(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}) \in \omega(\dot{x} + x, x, y)$. According to Lemma 8, it follows that

$$\lim_{t \rightarrow +\infty} H(\dot{x}(t) + x(t), x(t), y(t)) = H\left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}\right).$$

We consider the following two cases.

I. There exists $\bar{t} \geq 0$ such that

$$H(\dot{x}(\bar{t}) + x(\bar{t}), x(\bar{t}), y(\bar{t})) = H\left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}\right).$$

Since from Lemma 8(H1) we have

$$\frac{d}{dt} H(\dot{x}(t) + x(t), x(t), y(t)) \leq 0 \quad \forall t \in [0, +\infty),$$

we obtain for every $t \geq \bar{t}$ that

$$H(\dot{x}(t) + x(t), x(t), y(t)) \leq H(\dot{x}(\bar{t}) + x(\bar{t}), x(\bar{t}), y(\bar{t})) = H\left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}\right).$$

Thus $H(\dot{x}(t) + x(t), x(t), y(t)) = H\left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}\right)$ for every $t \geq \bar{t}$. According to Lemma 8(H1), it follows that $\dot{x}(t) = \dot{y}(t) = 0$ for almost every $t \in [\bar{t}, +\infty)$, hence x and y are constant on $[\bar{t}, +\infty)$ and the conclusion follows.

II. For every $t \geq 0$ it holds $H(\dot{x}(t) + x(t), x(t), y(t)) > H\left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}\right)$. Take $\Omega := \omega(\dot{x} + x, x, y)$.

By using Lemma 9(c) and (d) and the fact that H is a KL function, by Lemma 1, there exist positive numbers ϵ and η and a concave function $\varphi \in \Theta_\eta$ such that for all

$$\begin{aligned} (u, v, w) \in \{ & (u, v, w) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n : \text{dist}((u, v, w), \Omega) < \epsilon \} \\ & \cap \left\{ (u, v, w) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n : H\left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}\right) < H(u, v, w) < H\left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}\right) + \eta \right\}, \end{aligned} \quad (29)$$

one has

$$\varphi' \left(H(u, v, w) - H \left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x} \right) \right) \text{dist} \left((0, 0, 0), \partial H(u, v, w) \right) \geq 1. \quad (30)$$

Let $t_1 \geq 0$ be such that $H(\dot{x}(t) + x(t), x(t), y(t)) < H(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}) + \eta$ for all $t \geq t_1$. Since $\lim_{t \rightarrow +\infty} \text{dist} \left((\dot{x}(t) + x(t), x(t), y(t)), \Omega \right) = 0$, there exists $t_2 \geq 0$ such that for all $t \geq t_2$ the inequality $\text{dist} \left((\dot{x}(t) + x(t), x(t), y(t)), \Omega \right) < \epsilon$ holds. Hence for all $t \geq T := \max\{t_1, t_2\}$, $(\dot{x}(t) + x(t), x(t), y(t))$ belongs to the intersection in (29). Thus, according to (30), for every $t \geq T$ we have

$$\varphi' \left(H(\dot{x}(t) + x(t), x(t), y(t)) - H \left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x} \right) \right) \text{dist} \left((0, 0, 0), \partial H(\dot{x}(t) + x(t), x(t), y(t)) \right) \geq 1. \quad (31)$$

By applying Lemma 8(H2) we obtain for almost every $t \in [T, +\infty)$

$$\left(C_1 \|\dot{x}(t)\| + C_2 \|\dot{y}(t)\| \right) \varphi' \left(H(\dot{x}(t) + x(t), x(t), y(t)) - H \left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x} \right) \right) \geq 1, \quad (32)$$

where

$$C_1 := L + \frac{1}{\gamma} \text{ and } C_2 := \frac{2}{\gamma} + \frac{b}{\gamma a}.$$

From here, by using Lemma 8(H1), that $\varphi' > 0$ and

$$\begin{aligned} & \frac{d}{dt} \varphi \left(H(\dot{x}(t) + x(t), x(t), y(t)) - H \left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x} \right) \right) = \\ & \varphi' \left(H(\dot{x}(t) + x(t), x(t), y(t)) - H \left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x} \right) \right) \frac{d}{dt} H(\dot{x}(t) + x(t), x(t), y(t)), \end{aligned}$$

we deduce that for almost every $t \in [T, +\infty)$ it holds

$$\frac{d}{dt} \varphi \left(H(\dot{x}(t) + x(t), x(t), y(t)) - H \left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x} \right) \right) \leq -\frac{M_1 \|\dot{x}(t)\|^2 + M_2 \|\dot{y}(t)\|^2}{C_1 \|\dot{x}(t)\| + C_2 \|\dot{y}(t)\|}. \quad (33)$$

Let be $\alpha > 0$ (which does not depend on t) such that

$$-\frac{M_1 \|\dot{x}(t)\|^2 + M_2 \|\dot{y}(t)\|^2}{C_1 \|\dot{x}(t)\| + C_2 \|\dot{y}(t)\|} \leq -\alpha \|\dot{x}(t)\| - \alpha \|\dot{y}(t)\| \quad \forall t \geq 0.$$

From (33) we derive the inequality

$$\frac{d}{dt} \varphi \left(H(\dot{x}(t) + x(t), x(t), y(t)) - H \left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x} \right) \right) \leq -\alpha \|\dot{x}(t)\| - \alpha \|\dot{y}(t)\|, \quad (34)$$

which holds for almost every $t \geq T$. Since φ is bounded from below, by integration it follows $\dot{x}, \dot{y} \in L^1([0, +\infty); \mathbb{R}^n)$. From here we obtain that $\lim_{t \rightarrow +\infty} x(t)$ exists and the conclusion follows from the results obtained in this section. \blacksquare

Since the class of semi-algebraic functions is closed under addition (see for example [29]) and $(u, v, w) \mapsto c\|u - v\|^2 + c'\|av + bw\|^2$ is semi-algebraic for $c, c' > 0$, we obtain the following direct consequence of the above theorem.

Corollary 12 *Suppose that $f + \Phi$ is bounded from below and the parameters a, b, γ and L satisfy (13). For $x_0, y_0 \in \mathbb{R}^n$, let $(x, y) \in C^1([0, +\infty), \mathbb{R}^n) \times C^2([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (12). Suppose that x is bounded and $f + \Phi$ is semi-algebraic. Then the following statements are true:*

- (a) $\dot{x}, \dot{y}, ax + by \in L^1([0, +\infty); \mathbb{R}^n)$ and $\lim_{t \rightarrow +\infty} \dot{x}(t) = \lim_{t \rightarrow +\infty} \dot{y}(t) = \lim_{t \rightarrow +\infty} (ax(t) + by(t)) = 0$;
- (b) *there exists $\bar{x} \in \text{crit}(f + \Phi)$ such that $\lim_{t \rightarrow +\infty} x(t) = \bar{x}$ and $\lim_{t \rightarrow +\infty} y(t) = -\frac{a}{b}\bar{x}$.*

3.2 Convergence rates

In this subsection we investigate the convergence rates of the trajectories generated by the dynamical system (12). When solving optimization problems involving KL functions, convergence rates have been proved to depend on the so-called Lojasiewicz exponent (see [12, 26, 40, 45]). The main result of this subsection refers to the KL functions which satisfy Definition 1 for $\varphi(s) = Cs^{1-\theta}$, where $C > 0$ and $\theta \in (0, 1)$. We recall the following definition considered in [12].

Definition 3 Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. The function f is said to have the Lojasiewicz property, if for every $\bar{x} \in \text{crit } f$ there exist $C, \varepsilon > 0$ and $\theta \in (0, 1)$ such that

$$|f(x) - f(\bar{x})|^\theta \leq C\|x^*\| \text{ for every } x \text{ fulfilling } \|x - \bar{x}\| < \varepsilon \text{ and every } x^* \in \partial f(x). \quad (35)$$

According to [14, Lemma 2.1 and Remark 3.2(b)], the KL property is automatically satisfied at any noncritical point, fact which motivates the restriction to critical points in the above definition. The real number θ in the above definition is called *Lojasiewicz exponent* of the function f at the critical point \bar{x} .

Theorem 13 Suppose that $f + \Phi$ is bounded from below and the parameters a, b, γ and L satisfy (13). For $x_0, y_0 \in \mathbb{R}^n$, let $(x, y) \in C^1([0, +\infty), \mathbb{R}^n) \times C^2([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (12). Consider the function

$$H : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}, H(u, v, w) = (f + \Phi)(u) + \frac{1}{2\gamma}\|u - v\|^2 + \frac{1}{2\gamma a}\|av + bw\|^2.$$

Suppose that x is bounded and H satisfies Definition 1 for $\varphi(s) = Cs^{1-\theta}$, where $C > 0$ and $\theta \in (0, 1)$. Then there exists $\bar{x} \in \text{crit}(f + \Phi)$ such that $\lim_{t \rightarrow +\infty} x(t) = \bar{x}$ and $\lim_{t \rightarrow +\infty} y(t) = -\frac{a}{b}\bar{x}$. Let θ be the Lojasiewicz exponent of H at $(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}) \in \text{crit } H$, according to the Definition 3. Then there exist $a_1, b_1, a_2, b_2 > 0$ and $t_0 \geq 0$ such that for every $t \geq t_0$ the following statements are true:

- (a) if $\theta \in (0, \frac{1}{2})$, then x and y converge in finite time;
- (b) if $\theta = \frac{1}{2}$, then $\|x(t) - \bar{x}\| + \|y(t) + \frac{a}{b}\bar{x}\| \leq a_1 \exp(-b_1 t)$;
- (c) if $\theta \in (\frac{1}{2}, 1)$, then $\|x(t) - \bar{x}\| + \|y(t) + \frac{a}{b}\bar{x}\| \leq (a_2 t + b_2)^{-\left(\frac{1-\theta}{2\theta-1}\right)}$.

Proof. We define $\sigma : [0, +\infty) \rightarrow [0, +\infty)$ by (see also [26])

$$\sigma(t) = \int_t^{+\infty} \|\dot{x}(s)\| ds + \int_t^{+\infty} \|\dot{y}(s)\| ds \text{ for all } t \geq 0.$$

It is immediate that

$$\|x(t) - \bar{x}\| \leq \int_t^{+\infty} \|\dot{x}(s)\| ds \quad \forall t \geq 0. \quad (36)$$

Indeed, this follows by noticing that for $T \geq t$

$$\begin{aligned} \|x(t) - \bar{x}\| &= \|x(T) - \bar{x} - \int_t^T \dot{x}(s) ds\| \\ &\leq \|x(T) - \bar{x}\| + \int_t^T \|\dot{x}(s)\| ds, \end{aligned}$$

and by letting afterwards $T \rightarrow +\infty$.

Similarly we have

$$\left\| y(t) + \frac{a}{b}\bar{x} \right\| \leq \int_t^{+\infty} \|\dot{y}(s)\| ds \quad \forall t \geq 0. \quad (37)$$

From (36) and (37) we derive

$$\|x(t) - \bar{x}\| + \left\| y(t) + \frac{a}{b}\bar{x} \right\| \leq \sigma(t) \quad \forall t \geq 0. \quad (38)$$

We assume that for every $t \geq 0$ we have $H(\dot{x}(t) + x(t), x(t), y(t)) > H(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x})$. As seen in the proof of Theorem 11, in the other case the conclusion follows automatically. Furthermore, by invoking again the proof of above-named result, there exist $t_0 \geq 0$ and $\alpha > 0$ such that for almost every $t \geq t_0$ (see (33))

$$\alpha \|\dot{x}(t)\| + \alpha \|\dot{y}(t)\| + \frac{d}{dt} \left[H(\dot{x}(t) + x(t), x(t), y(t)) - H\left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}\right) \right]^{1-\theta} \leq 0$$

and

$$\left\| (\dot{x}(t) + x(t), x(t), y(t)) - \left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}\right) \right\| < \varepsilon.$$

We derive by integration (for $T \geq t \geq t_0$)

$$\begin{aligned} \alpha \int_t^T \|\dot{x}(s)\| ds + \alpha \int_t^T \|\dot{y}(s)\| ds + \left[H(\dot{x}(T) + x(T), x(T), y(T)) - H\left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}\right) \right]^{1-\theta} \\ \leq \left[H(\dot{x}(t) + x(t), x(t), y(t)) - H\left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}\right) \right]^{1-\theta}, \end{aligned}$$

hence

$$\alpha \sigma(t) \leq \left[H(\dot{x}(t) + x(t), x(t), y(t)) - H\left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}\right) \right]^{1-\theta} \quad \forall t \geq t_0. \quad (39)$$

Since θ is the Łojasiewicz exponent of H at $(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x})$, we have

$$\left| H(\dot{x}(t) + x(t), x(t), y(t)) - H\left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}\right) \right|^\theta \leq C \|x^*\| \quad \forall x^* \in \partial H(\dot{x}(t) + x(t), x(t), y(t))$$

for every $t \geq t_0$. According to Lemma 8(H2), we can find $x^*(t) \in \partial H(\dot{x}(t) + x(t), x(t), y(t))$ and a constant $N > 0$ such that for every $t \in [0, +\infty)$

$$\|x^*(t)\| \leq N \|\dot{x}(t)\| + N \|\dot{y}(t)\|.$$

From the above two inequalities we derive for almost every $t \in [t_0, +\infty)$

$$\left| H(\dot{x}(t) + x(t), x(t), y(t)) - H\left(\bar{x}, \bar{x}, -\frac{a}{b}\bar{x}\right) \right|^\theta \leq C \cdot N \|\dot{x}(t)\| + C \cdot N \|\dot{y}(t)\|,$$

which combined with (39) yields

$$\alpha \sigma(t) \leq (C \cdot N \|\dot{x}(t)\| + C \cdot N \|\dot{y}(t)\|)^{\frac{1-\theta}{\theta}}. \quad (40)$$

Since

$$\dot{\sigma}(t) = -\|\dot{x}(t)\| - \|\dot{y}(t)\| \quad (41)$$

we conclude that there exists $\alpha' > 0$ such that for almost every $t \in [t_0, +\infty)$

$$\dot{\sigma}(t) \leq -\alpha' (\sigma(t))^{\frac{\theta}{1-\theta}}. \quad (42)$$

If $\theta = \frac{1}{2}$, then

$$\dot{\sigma}(t) \leq -\alpha' \sigma(t)$$

for almost every $t \in [t_0, +\infty)$. By multiplying with $\exp(\alpha't)$ and integrating afterwards from t_0 to t , it follows that there exist $a_1, b_1 > 0$ such that

$$\sigma(t) \leq a_1 \exp(-b_1 t) \quad \forall t \geq t_0$$

and the conclusion of (b) is immediate from (38).

Assume that $0 < \theta < \frac{1}{2}$. We obtain from (42)

$$\frac{d}{dt} \left(\sigma(t)^{\frac{1-2\theta}{1-\theta}} \right) \leq -\alpha' \frac{1-2\theta}{1-\theta}$$

for almost every $t \in [t_0, +\infty)$.

By integration we get

$$\sigma(t)^{\frac{1-2\theta}{1-\theta}} \leq -\bar{\alpha}t + \bar{\beta} \quad \forall t \geq t_0,$$

where $\bar{\alpha} > 0$. Thus there exists $T \geq 0$ such that

$$\sigma(T) \leq 0 \quad \forall t \geq T,$$

which implies that x and y are constant on $[T, +\infty)$.

Finally, suppose that $\frac{1}{2} < \theta < 1$. We obtain from (42)

$$\frac{d}{dt} \left(\sigma(t)^{\frac{1-2\theta}{1-\theta}} \right) \geq \alpha' \frac{2\theta-1}{1-\theta}$$

for almost every $t \in [t_0, +\infty)$. By integration one derives

$$\sigma(t) \leq (a_2 t + b_2)^{-\left(\frac{1-\theta}{2\theta-1}\right)} \quad \forall t \geq t_0,$$

where $a_2, b_2 > 0$. Statement (c) follows from (38). ■

References

- [1] B. Abbas, H. Attouch, *Dynamical systems and forward-backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator*, Optimization 64(10), 2223–2252, 2015
- [2] B. Abbas, H. Attouch, B.F. Svaiter, *Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces*, Journal of Optimization Theory and its Applications 161(2), 331–360, 2014
- [3] F. Alvarez, *On the minimizing property of a second order dissipative system in Hilbert spaces*, SIAM Journal on Control and Optimization 38(4), 1102–1119, 2000
- [4] F. Alvarez, *Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space*, SIAM Journal on Optimization 14(3), 773–782, 2004
- [5] F. Alvarez, H. Attouch, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Analysis 9(1-2), 311, 2001

- [6] F. Alvarez, H. Attouch, J. Bolte, P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics*, Journal de Mathématiques Pures et Appliquées (9) 81(8), 747–779, 2002
- [7] F. Alvarez, J.M. Pérez, *A dynamical system associated with Newton’s method for parametric approximations of convex minimization problems*, Applied Mathematics and Optimization 38(2), 193–217, 1998
- [8] A.S. Antipin, *Minimization of convex functions on convex sets by means of differential equations*, (Russian) Differentsial’nye Uravneniya 30(9), 1475–1486, 1994; translation in Differential Equations 30(9), 1365–1375, 1994
- [9] H. Attouch, F. Alvarez, *The heavy ball with friction dynamical system for convex constrained minimization problems*, in: Optimization (Namur, 1998), 25–35, in: Lecture Notes in Economics and Mathematical Systems 481, Springer, Berlin, 2000
- [10] H. Attouch, G. Buttazzo, G. Michaille, *Variational Analysis in Sobolev and BV Spaces: Applications to PDEs and Optimization, Second Edition*, MOS-SIAM Series on Optimization, Philadelphia, 2014
- [11] H. Attouch, M. Marques Alves, B.F. Svaiter, *A dynamic approach to a proximal-Newton method for monotone inclusions in Hilbert spaces, with complexity $O(1/n^2)$* , Journal of Convex Analysis 23(1), 139–180, 2016
- [12] H. Attouch, J. Bolte, *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*, Mathematical Programming 116(1-2) Series B, 5–16, 2009
- [13] H. Attouch, J. Bolte, P. Redont, *Optimizing properties of an inertial dynamical system with geometric damping. Link with proximal methods*, Control and Cybernetics 31(3), 643–657, 2002
- [14] H. Attouch, J. Bolte, P. Redont, A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality*, Mathematics of Operations Research 35(2), 438–457, 2010
- [15] H. Attouch, J. Bolte, B.F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods*, Mathematical Programming 137(1-2) Series A, 91–129, 2013
- [16] H. Attouch, M.-O. Czarnecki, *Asymptotic behavior of coupled dynamical systems with multiscale aspects*, Journal of Differential Equations 248(6), 1315–1344, 2010
- [17] H. Attouch, X. Goudou, P. Redont, *The heavy ball with friction method. I. The continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system*, Communications in Contemporary Mathematics 2(1), 1–34, 2000
- [18] H. Attouch, P.-E. Maingé, P. Redont, *A second-order differential system with Hessian-driven damping; application to non-elastic shock laws*, Differential Equations and Applications 4(1), 27–65, 2012
- [19] H. Attouch, J. Peypouquet, P. Redont, *Fast convex optimization via inertial dynamics with Hessian driven damping*, Journal of Differential Equations 261, 5734–5783, 2016
- [20] H. Attouch, P. Redont, *The second-order in time continuous Newton method*, in Approximation, optimization and mathematical economics (Pointe-à-Pitre, 1999), 25–36, Physica, Heidelberg, 2001

- [21] H. Attouch, B.F. Svaiter, *A continuous dynamical Newton-like approach to solving monotone inclusions*, SIAM Journal on Control and Optimization 49(2), 574–598, 2011
- [22] S. Banert, R.I. Boţ, *A forward-backward-forward differential equation and its asymptotic properties*, to appear in Journal of Convex Analysis, arXiv:1503.07728, 2015
- [23] H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, New York, 2011
- [24] H.H. Bauschke, P.L. Combettes, D.R. Luke, *Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization*, Journal of the Optical Society of America A. Optics, Image Science, and Vision 19(7), 1334–1345, 2002
- [25] J. Bolte, *Continuous gradient projection method in Hilbert spaces*, Journal of Optimization Theory and its Applications 119(2), 235–259, 2003
- [26] J. Bolte, A. Daniilidis, A. Lewis, *The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM Journal on Optimization 17(4), 1205–1223, 2006
- [27] J. Bolte, A. Daniilidis, A. Lewis, M. Shota, *Clarke subgradients of stratifiable functions*, SIAM Journal on Optimization 18(2), 556–572, 2007
- [28] J. Bolte, A. Daniilidis, O. Ley, L. Mazet, *Characterizations of Lojasiewicz inequalities: subgradient flows, talweg, convexity*, Transactions of the American Mathematical Society 362(6), 3319–3363, 2010
- [29] J. Bolte, S. Sabach, M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Mathematical Programming Series A (146)(1–2), 459–494, 2014
- [30] J.M. Borwein, Q.J. Zhu, *Techniques of Variational Analysis*, Springer, New York, 2005
- [31] R.I. Boţ, E.R. Csetnek, *An inertial Tseng’s type proximal algorithm for nonsmooth and nonconvex optimization problems*, Journal of Optimization Theory and Applications, DOI 10.1007/s10957-015-0730-z
- [32] R.I. Boţ, E.R. Csetnek, *A dynamical system associated with the fixed points set of a nonexpansive operator*, Journal of Dynamics and Differential Equations, DOI: 10.1007/s10884-015-9438-x, 2015
- [33] R.I. Boţ, E.R. Csetnek, *A second order dynamical system with Hessian-driven damping and penalty term associated to variational inequalities*, arXiv:1608.04137, 2016
- [34] R.I. Boţ, E.R. Csetnek, *Approaching the solving of constrained variational inequalities via penalty term-based dynamical systems*, Journal of Mathematical Analysis and Applications 435(2), 1688–1700, 2016
- [35] R.I. Boţ, E.R. Csetnek, *Second order forward-backward dynamical systems for monotone inclusion problems*, Siam Journal on Control and Optimization 54(3), 1423–1443, 2016
- [36] R.I. Boţ, E.R. Csetnek, *Convergence rates for forward-backward dynamical systems associated with strongly monotone inclusions*, to appear in Journal of Mathematical Analysis and Applications, arXiv:1504.01863, 2015
- [37] R.I. Boţ, E.R. Csetnek, S. László, *An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions*, EURO Journal on Computational Optimization 4, 3–25, 2016

- [38] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Mathematics Studies No. 5, Notas de Matemática (50), North-Holland/Elsevier, New York, 1973
- [39] E. Chouzenoux, J.-C. Pesquet, A. Repetti, *Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function*, Journal of Optimization Theory and its Applications 162(1), 107–132, 2014
- [40] P. Frankel, G. Garrigos, J. Peypouquet, *Splitting methods with variable metric for Kurdyka-Lojasiewicz functions and general convergence rates*, Journal of Optimization Theory and its Applications 165(3), 874–900, 2015
- [41] A. Haraux, *Systèmes Dynamiques Dissipatifs et Applications*, Recherches en Mathématiques Appliquées 17, Masson, Paris, 1991
- [42] A. Haraux, M. Jendoubi, *Convergence of solutions of second-order gradient-like systems with analytic nonlinearities*, Journal of Differential Equations 144(2), 313–320, 1998
- [43] R. Hesse, D.R. Luke, S. Sabach, M.K. Tam, *Proximal heterogeneous block input-output method and application to blind ptychographic diffraction imaging*, SIAM Journal on Imaging Sciences 8(1), 426–457, 2015
- [44] K. Kurdyka, *On gradients of functions definable in o-minimal structures*, Annales de l’institut Fourier (Grenoble) 48(3), 769–783, 1998
- [45] S. Łojasiewicz, *Une propriété topologique des sous-ensembles analytiques réels*, Les Équations aux Dérivées Partielles, Éditions du Centre National de la Recherche Scientifique Paris, 87–89, 1963
- [46] B. Mordukhovich, *Variational Analysis and Generalized Differentiation, I: Basic Theory, II: Applications*, Springer-Verlag, Berlin, 2006
- [47] P. Ochs, Y. Chen, T. Brox, T. Pock, *iPiano: Inertial proximal algorithm for non-convex optimization*, SIAM Journal on Imaging Sciences 7(2), 1388–1419, 2014
- [48] R.T. Rockafellar, R.J.-B. Wets, *Variational Analysis*, Fundamental Principles of Mathematical Sciences 317, Springer-Verlag, Berlin, 1998
- [49] L. Simon, *Asymptotics for a class of nonlinear evolution equations, with applications to geometric problems*, Annals of Mathematics (2) 118, 525–571, 1983